Motifs in Triadic Random Graphs based on Steiner Triple Systems

Marco Winkler[1, *] and Jörg Reichardt[1, †]

[1]*Institute for Theoretical Physics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany*

Conventionally, pairwise relationships between nodes are considered to be the fundamental building blocks of complex networks. However, over the last decade the overabundance of certain sub-network patterns, so called motifs, has attracted high attention. It has been hypothesized, these motifs, instead of links, serve as the building blocks of network structures.

Although the relation between a network's topology and the general properties of the system, such as its function, its robustness against perturbations, or its efficiency in spreading information is the central theme of network science, there is still a lack of sound generative models needed for testing the functional role of subgraph motifs. Our work aims to overcome this limitation.

We employ the framework of exponential random graphs (ERGMs) to define novel models based on triadic substructures. The fact that only a small portion of triads can actually be set independently poses a challenge for the formulation of such models. To overcome this obstacle we use Steiner Triple Systems (STS). These are partitions of sets of nodes into pair-disjoint triads, which thus can be specified independently. Combining the concepts of ERGMs and STS, we suggest novel generative models capable of generating ensembles of networks with non-trivial triadic Z-score profiles. Further, we discover inevitable correlations between the abundance of triad patterns, which occur solely for statistical reasons and need to be taken into account when discussing the functional implications of motif statistics. Moreover, we calculate the degree distributions of our triadic random graphs analytically.

I. INTRODUCTION

The topological structure of interactions among the constituents of complex many particle systems is intimately linked to system function and global system properties. The study of complex networks aims to elucidate this link between structure and function.

Motivated by the stark contrast between topological features found in real-world data and expectations based on the assumption of purely random link formation [1, 2], two main threads of research can be identified.

The first thread is aimed predominantly at explaining the network formation process, *i.e.* identifying the forces shaping a network. A particularly productive approach has been the development of network growth models following the publication of Barabási and Albert [3] to explain non-Poissonian degree distibutions. See [4–6] and the references therein for a review. Growth models generally take the agreement between a particular feature in real-world data with networks resulting from a particular model as evidence for a particular aspect of a growth process, such as preferential attachment.

The second thread of research focusses on explaining the influence certain topological features may have on global system properties such as the robustness against perturbations or the stability of the system under node or link removal [7], as well as on dynamical processes taking place on the network [8]. In order to study such questions systematically, the ability to generate an ensemble of networks with a precise set of topological fea-

tures, but no others, is crucial. Growth models are generally not suited for this task since a network formation process often introduces invariable correlations between network features that are difficult to disentangle. For example, the Barabási-Albert model is capable of generating networks with a broad degree distribution, but at the same time introduces degree-degree correlations. Further, growth models are generally very difficult to characterize in terms of their statistical properties. In contrast, generative probabilistic models which parameterize an ensemble of networks via an explicit expression for the probability distribution over adjacency matrices can facilitate such analysis. The present work introduces a new class of probabilistic generative models.

Good generative probabilistic models of networks should combine three characteristics: First, every aspect of network structure that is not explicitly specified through the parametrization is maximally random. Second, they should allow for unbiased estimation of parameters from data. If parameters are estimated from data, these data are typical for the ensemble thus parameterized. Third, they should be easy to specify and parameters should be simple to learn and interpret. Exponential random graph models (ERGMs), *i.e.* those that specify a Boltzmann distribution over the set of all adjacency matrices of given size, meet all of these criteria [9–12]. They are maximum entropy, mean unbiased and parameters can be learnt consistently via maximum likelihood estimators or Monte Carlo Markov Chain (MCMC) methods [13–15].

Generally, pairwise relations between nodes, so-called dyads, are considered the fundamental building blocks of complex networks and hence also the fundamental unit when modeling a network, regardless whether by growth

* mwinkler@physik.uni-wuerzburg.de
† reichardt@physik.uni-wuerzburg.de
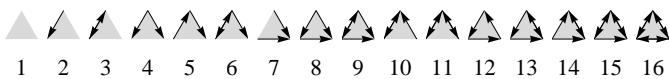
FIG. 1. All 16 possible non-isomorphic triadic subgraphs (subgraph patterns) in directed networks.

or probabilistic models. Erdös-Rényi (ER) graphs, the configuration model [16, 17], stochastic block models [18–21] and degree corrected block models [22] all fall into the class of dyadic models. The basic assumption underlying dyadic models is that dyads are *conditionally independent* given the model's parameters.

However, the assumption of dyadic independence as a general paradigm of network modeling seems questionable. For example, in a social context, the idea that the relation of Alice and Bob be independent from the relation of Alice and Charlie seems to go against experience, especially if the relation is of romantic type. Similarly, triadic closure, or the large clustering coefficient observed in many networks, hints at a dependence between the connections in a network. Generalizing these ideas, during the last decade the systematic study of third and fourth order sub-network structure captured high attention [23–27]. Apart from node permutations, there are 16 distinct triad patterns in directed unweighted networks as shown in Fig. 1. It was found that certain patterns of three-node subgraphs occur significantly more frequent than expected in an ensemble of networks generated by shuffling the connections of the original network under the constraint of preserving the nodes' in- and out-degrees.

Sampling from an ensemble of randomized networks yields an average occurrence $\langle N_{\mathrm{rand},i} \rangle$ and a standard deviation $\sigma_{\mathrm{rand},i}$ for each triad pattern i shown in Fig. 1. Over- and underrepresentation of pattern i is quantified by through a Z-score

$$Z_i = \frac{N_{\mathrm{original},i} - \langle N_{\mathrm{rand},i} \rangle}{\sigma_{\mathrm{rand},i}}. \tag{1}$$

Notice that Z-scores are evaluated by counting the subgraph patterns over all $\binom{N}{3}$ possible triads. Every network can be assigned a vector $\vec{Z}$ whose components comprise the Z-scores of all possible triad patterns. Significant patterns are referred to as 'motifs' [23]. It is common to consider only the Z-scores of the triad patterns in which all three nodes are attached to an edge. Further, one commonly refers to the normalized Z-vector as the 'significance profile', $\vec{\mathrm{SP}} = \vec{Z}/\sqrt{\sum_{i=1}^{13} Z_i^2}$. This normalization makes systems of different sizes comparable [23]. Many real-world systems have been examined with respect to their triadic Z-scores and significance profiles [23, 24, 28–30] and it was suggested that they can be grouped into so-called 'super-families' [24].

Surprisingly, to date, no general model exists that can fully explain or model the triad significance profiles observed in many real world networks. The present work suggests a generative probabilistic model capable of describing a wide range of significance profiles.

A number of growth models exist which are capable of reproducing certain parts of the motif statistics, in particular, the fraction of closed triangles by explicitly formulating 'triadic closure' processes. Starting from an initially unclustered network, one searches for edges with a common neighbor and then connects them successively to form triangles [31–36]. Yet, the calculation of their properties is limited to numerical approaches [31].

Further, specifying generative models has proven difficult. Using the Strauss model [10], specified by a Hamiltonian with two fields, one acting on individual links, the other one acting on triads of links, it is possible to generate systems with - on average - predefined link and triad appearance. However, Park and Newman could show that the average does not describe the properties of a typical system generated by the model. In fact, there is a large degenerate phase in which most instances of networks tend to be either fully connected or empty [37].

Another alternative suggested by Newman generates networks in which both the number of single links, s_i, of every node i, as well as the number of triads, t_i, it participates in are specified initially [31]. The model yields networks, drawn uniformly at random from the set of all possible matchings of 'stubs' and 'corners'. With this generalization of random-graph models, it is possible to compute analytically component sizes, the existence and size of a giant component, and percolation properties. The model yields an unbiased ensemble of networks with clustering. However, attempting to specify the probabilities for all possible three-node subgraphs simultaneously poses a problem.

Alternatively, it has been noted early on that latent variables might offer an explanation for the observe motif distributions within the framework of dyadic independence models. The randomization employed in the calculation of the Z-scores ignores all mesoscopic structure, possibly present in the system. Thus, parts of the over- and underrepresentations of certain motifs, compared to the randomized versions, may stem from such structure [38–40]. E.g. some features of the significance profile of the neural network of *C. elegans* could successfully be explained by means of latent class structure, while accounting for both properties on the individual node level and on the group level [38]. The abundance of triad motifs is apparently strongly related to mesoscopic network structure or, in other words, comparison of a network with block structure to a null model which does not account for such groups may result in Z-scores which are more than less artifacts of the mesoscopic structure. Yet, mesoscopic block models alone are not sufficient to explain all observed motifs.

In general, when trying to reproduce triad structures, models formulated in terms of dyads face the difficulty that each dyad influences an extensive number of triads. On the other hand, directly modeling all triad structures

is impossible, as not all local triad configurations may be specified independently from each other. Yet, the Z-score statistics are obtained by considering every individual triadic subgraph pattern.

In the following section we will suggest a model which is based on triads rather than dyads which actually *can* be specified independently from each other, so-called Steiner Triple Systems (STSs). Starting from the framework of Steiner Triple Systems, it will be possible to define a whole class of triadic exponential random graph models. In this paper we discuss the most basic of such models, one which assumes the same probability distribution of triadic subgraph configurations on all Steiner Triples (STs). This can be considered the triadic analogon to ER graphs on dyadic models. We will investigate how a distribution on the STs affects the correpsonding triad significance profiles. With this work, we will be able to investigate correlations in the abundance of triad patterns which occur solely for statistical reasons. Moreover, we provide for a class of generative models which are capable of modeling structure of higher than dyadic order. We aim to design ensembles of networks with pre-defined Z-score profiles. In section II we will introduce the concept of STSs, subsequently in section III we will define the triadic random graph model, a generative model based on STs. Finally, in section IV we will present results for the latter. In particular, we will show that triadic random graphs are capable of generating networks with non-vanishing Z-scores. Furthermore, we will investigate correlations in the appearance of triadic subgraph patterns and discuss their implications for the functional interpretation of motif significance profiles. Finally, we will calculate the degree distribution of triadic random graphs analytically.

II. STEINER TRIPLE SYSTEMS

We will now define the terminology used throughout the remainder of this article: A *dyad* is a set of two nodes. An *edge*, or interchangeably a *link*, describes the presence of a dyadic connection, *i.e.* a connection between two nodes; it can be uni- or bidirectional. A *triad* is a set of three nodes. A *triangle* denotes three mutually interconnected nodes. A *subgraph* is a part of a network which considers only a subset of all nodes, including their mutual connections. A *subgraph configuration* is a specification of the connections in a subgraph, while accounting for node identities; e.g. dyad configuration $A \rightarrow B$ is distinct from dyad configuration $B \leftarrow A$. *Subgraph patterns* are sets of nodes including their relations without accounting for node identities, *i.e.* isomorphic subgraph configurations are mapped to the same subgraph pattern; e.g. dyad pattern $A \rightarrow B$ is the same as pattern $A \leftarrow B$. A(n) *(anti-)motif* is a subgraph pattern which is significantly over-(under-) represented, as compared to some null model.

In a network of N nodes there are $T = \binom{N}{3}$ distinct
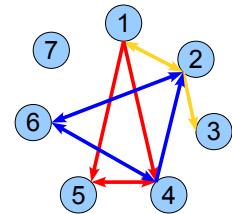
FIG. 2. (Color online) Only few triad configurations can be specified independently of each other: e.g. a specification of the triads $(1, 2, 3)$, $(1, 4, 5)$, and $(2, 4, 6)$ fully determines the configuration of $(1, 2, 4)$.

triads. Yet, it is not possible to specify all their triadic-subgraph configurations independently of each other; e.g. consider the network in Fig. 2. Suppose we set the relations in the three-node subgraph of nodes 1, 2, and 3, denoted as $(1, 2, 3)$, such that they adopt pattern �₳. Further, we specify the triads $(1, 4, 5)$ and $(4, 6, 2)$ such that they assume patterns ◬ and ◺, respectively. Then, with the choices for the discussed three triads in Fig. 2, the subgraph of $(4, 1, 2)$ is already determined to take the pattern ◬ implicitly. This is because $(4, 1, 2)$ contains dyadic relations which have already been assigned in the other three triads.

Since there are only $E = \binom{N}{2}$ dyads in a network and every triad comprises three dyadic relations, there is an upper bound to the number of triads which are dyad-disjoint and therefore can be set without over-determining the system:

$$\# \text{ of dyad-disjoint triads} \leq \frac{E}{3} = \frac{N(N-1)}{6} \ll T \quad (2)$$

Networks for which the upper bound is exactly met can be partitioned into triads such that every pair of nodes in the system is part of *exactly one* of them. Such systems are called Steiner Triple Systems (STS) [41]. STSs consisting of N vertices are called Steiner Triple Systems of order N, or STS(N). There are two neccessary and sufficient requirements for the existence of an STS(N):

$$N \bmod 2 = 1 \quad (3)$$
$$N(N-1) \bmod 3 = 0 \quad (4)$$

For a detailed discussion see e.g. [42, page 277ff] or [43, page 205ff]. The problem was originally solved by Kirkman in 1847 [44].

From Eq. (3) and Eq. (4) we can conclude that, by approximation, systems of arbitrary size can be decomposed into Steiner Triples. All one has to do is either add up to three 'dummy' nodes to the system, or to ignore up to three nodes including their relations.

To fix ideas, Fig. 3 shows the partition of a STS of order 7 into STs. Due to the small amount of vertices it is possible to derive the STS deductively: Without loss of generality, we start with node 1. Since 1 is part of six

FIG. 3. (Color online) Schematic presentation of a Steiner Triple System of order seven. The Steiner Triples are set to be (1,2,3), (1,4,5), (1,6,7), (2,4,6), (2,5,7), (3,4,7), and (3,5,6), as indicated by the colors of the matrix elements. Every matrix element is assigned to exactly one Steiner Triple.

(1,2,3)	(1,3,6)	**(1,6,7)**	(2,4,7)	**(3,5,6)**
(1,2,4)	(1,3,7)	(2,3,4)	(2,5,6)	(3,5,7)
(1,2,5)	**(1,4,5)**	(2,3,5)	**(2,5,7)**	(3,6,7)
(1,2,6)	(1,4,6)	(2,3,6)	(2,6,7)	(4,5,6)
(1,2,7)	(1,4,7)	(2,3,7)	(3,4,5)	(4,5,7)
(1,3,4)	(1,5,6)	(2,4,5)	(3,4,6)	(4,6,7)
(1,3,5)	(1,5,7)	**(2,4,6)**	**(3,4,7)**	(5,6,7)

TABLE I. Possible specification of Steiner Triples for a system of size 7.

dyads, one with every remaining node, it has to be part of three Steiner Triples. The first one shall be (1,2,3) (color coded in yellow in Fig. 3), the second one (1,4,5) (red), and the third one (1,6,7) (cyan). Now each dyadic relation 1 participates in is covered by exactly one Steiner Triple. We continue with the dyads of node 2: those with nodes 1 and 3 are already contained in $(1,2,3)$. 4 and 5 are already part of ST $(1,4,5)$ and therefore need to be assigned to distinct ST. We choose 6 to be in the ST with 2 and 4 (blue), and thus we have also specified ST $(1,5,7)$ (green). Continuing with node 3, the dyads with nodes 4, 5, 6, and 7 need to be assigned to ST. 4 is already assigned to STs with 5 and 6. Thus, the two remaining ST are $(3,4,7)$ (magenta) and $(3,5,6)$ (orange).

From the $\binom{7}{3} = 35$ possible triads of a network of order seven only $E/3 = 7 \cdot 6/6 = 7$ triads can be specified independently from each other.

In table I all triads for a network of seven nodes are displayed. A possible choice of an STS is highlited with colors corresponding to Fig. 3.

A detailed proof that Eq. (3) and Eq. (4) are indeed sufficient for the existence of an STS can be found in [43].

Of course, for larger system sizes it is not practical to construct STSs the way described above. However, larger STS can be constructed by merging smaller ones. For the STS(7), the partition described above is unique, apart from relabeling nodes. For STSs of higher order, there are multiple non-isomorphic ways to partition the nodes into Steiner Triples. STSs provide us with sets of triads which can actually be configured without overdetermining dyadic relations. They thus, can be considered a basis to express an adjacency matrix.

In order to account for substructures of higher than dyadic order, our goal is now to define a model based on triadic rather than dyadic entities. Since Steiner Triple Systems assign every dyadic relation, *i.e.* every pair of nodes, to exactly one triad, the specification of the configurations of all Steiner Triples is equivalent to specifying an adjacency matrix $\boldsymbol{A}$. To convince oneself that a formulation of a network in terms of Steiner Triples is equivalent to a formulation in terms of dyads, consider a directed unweighted graph with N vertices. There are $\binom{N}{2}$ dyads. Each dyad (i,j) may adopt four distinct configurations. Thus, in total there are $4^{\binom{N}{2}} = 2^{2\binom{N}{2}}$ possible states of the system, *i.e.* distinct adjacency matrices. On the other hand, there are $\binom{N}{2}/3$ distinct Steiner Triples. Each of those triads may assume $2^6 = 64$ distinct configurations (every of the six unidirectional links in the triad may be present or absent). Therefore, again we obtain $64^{\binom{N}{2}/3} = 2^{6\binom{N}{2}/3} = 2^{2\binom{N}{2}}$ possible states. The argument for undirected graphs is analogous.

III. MODEL

Let us recall that dyadic ERGMs assume that the likelihoods for the presence of two edges are conditionally independent of each other. Further, let the matrix $\boldsymbol{D}$ with components $D_{ij} \in \{0,1\}$ denote the random variables corresponding to the entries of the adjacency matrix $\boldsymbol{A}$. Then, the independence assumption implies for the likelihood of observing an adjacency matrix $\boldsymbol{A}$

$$\mathcal{P}\left(\mathbf{D} = \mathbf{A} \,|\, \vec{\theta}\right) = \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \mathcal{P}\left(D_{ij} = A_{ij}, D_{ji} = A_{ji} | \vec{\theta}\right)$$

$$= \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \mathcal{P}\left(\vec{D}_{(i,j)} = \vec{A}_{(i,j)} | \vec{\theta}\right) \quad (5)$$

where $\vec{\theta}$ includes all parameters of the model. The vector notation on the right hand side accounts for the fact that in directed unweighted networks, there are four possible dyadic relations: $A_{ij} = 0 \wedge A_{ji} = 0$, $A_{ij} = 0 \wedge A_{ji} = 1$, $A_{ij} = 1 \wedge A_{ji} = 0$, and $A_{ij} = 1 \wedge A_{ji} = 1$. They can be combined in a four dimensional indicator vector $\vec{A}_{(i,j)}$ with all components being zero, except for one being one.

We will now employ the concept of Steiner Triple Systems to define the triadic analogon to Eq. (5). Now, instead of assuming the likelihoods of *dyads* to be conditionally independent of each other, we suppose the likelihoods for the configurations on *Steiner Triples* to be conditionally independent. With this asumption, the likelihood of observing an adjacency matrix $\boldsymbol{A}$ factorizes as follows:

$$\mathcal{P}\left(\mathbf{D} = \mathbf{A} \,|\, \vec{\theta}\right) = \prod_{\sigma=1}^{N(N-1)/6} \mathcal{P}\left(\vec{D}_\sigma = \vec{A}_\sigma | \vec{\theta}\right)$$

$$= \prod_{\sigma=1}^{N(N-1)/6} \vec{\mathcal{P}}\left(\vec{D}_\sigma | \vec{\theta}\right) \cdot \vec{A}_\sigma \quad (6)$$
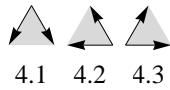
FIG. 4. The three isomorphic configurations belonging to pattern 4.

where σ denotes the Steiner Triples of an STS(N), $\vec{D}_\sigma$ is an indicator variable for the configuration of Steiner Triple σ, and $\vec{A}_\sigma$ is a value of this variable. Analogously to Eq. (5), for each of the vectors exactly one component is unity, while all others are zero, which is equivalent to the fact that a triad cannot be in multiple configurations at the same time. For undirected networks it is $\vec{D}_\sigma \in \{0,1\}^8$, for directed ones it is $\vec{D}_\sigma \in \{0,1\}^{64}$. Accordingly, it is $\vec{\mathcal{P}}\left(\vec{D}_\sigma|\vec{\theta}\right) \in [0,1]^8$ or $[0,1]^{64}$, respectively with the sums of the elements normalized to one. By defining Eq. (6) we make the assumption that the likelihoods of Steiner Triple configurations factorize, *i.e.* they are conditionally independent of each other.

For unweighted graphs, Eq. (6) describes the most general formulation of models based on conditionally independent STs. We will now further investigate the properties of a particular realization of this class of models. The simplest such model has the same likelihood distribution for the triad configurations on all Steiner Triples, $\vec{\mathcal{P}}\left(\vec{D}_\sigma|\vec{\theta}\right) = \vec{\mathcal{P}}\left(\vec{D}|\vec{\theta}\right)$. This can be regarded the triadic analogon to dyadic ER graphs, in which the likelihood for the existence of an edge is the same for all dyads. We will refer to them as *triadic random graphs*. Since the ordering of the nodes in a Steiner Triple is arbitrary there is no need to distinguish between isomorphic triad configurations. E.g. the likelihoods of the three configurations of subgraph 4, shown in Fig. 4, will be the same. Thus, the triadic random graphs have 16 parameters each of them indicating the probability of a Steiner Triple to assume one of the subgraphs shown in Fig. 1. Of course, their values need to sum up to unity.

Given the parameters, the probability disitribution of each Steiner Triple is given by:

$$
\begin{aligned}
\vec{\mathcal{P}}\left(\vec{D}|\vec{\theta}\right) &= \boldsymbol{M}\,\vec{\mathcal{P}} \\
&= \boldsymbol{M}\left(p(\triangle), p(\measuredangle), p(\measuredangle), p(\wedge), p(\wedge), p(\wedge), p(\triangle), p(\triangle), p(\triangle), p(\wedge), p(\wedge), p(\triangle), p(\triangle), p(\triangle), p(\triangle), p(\triangle)\right)^T
\end{aligned}
\tag{7}
$$

The matrix $\boldsymbol{M}$ maps each of the 16 non-isomorphic subgraph patterns in Fig. 1 to their corresponding isomorphic configurations with equal probability, *i.e.* the sums of its columns are normalized to one. Here, the only parameters $\vec{\theta}$ of the model are the entries of the vector $\vec{\mathcal{P}}$.

Eq. (6) and Eq. (7) describe the triadic random graph model, in which the configuration for each Steiner Triple is drawn - conditionally independent from other Steiner Triples - from the same probability distribution over the 16 subgraphs shown in Fig. 1.

If (unidirectional) links are set purely at random with probability p as it is the case in ER graphs, the probabilities for the triadic subgraph patterns are:

$$
\begin{aligned}
p_{\triangle}^{ER} &= (1-p)^6 \\
p_{\measuredangle}^{ER} &= 6\,p\,(1-p)^5 \\
p_{\measuredangle}^{ER} &= p_{\wedge}^{ER} = p_{\triangle}^{ER} = 3\,p^2\,(1-p)^4 \;,\; p_{\wedge}^{ER} = 6\,p^2\,(1-p)^4 \\
p_{\wedge}^{ER} &= p_{\wedge}^{ER} = p_{\triangle}^{ER} = 6\,p^3\,(1-p)^3 \;,\; p_{\triangle}^{ER} = 2\,p^3\,(1-p)^3 \\
p_{\triangle}^{ER} &= p_{\triangle}^{ER} = p_{\triangle}^{ER} = 3\,p^4\,(1-p)^2 \;,\; p_{\triangle}^{ER} = 6\,p^4\,(1-p)^2 \\
p_{\triangle}^{ER} &= 6\,p^5\,(1-p) \\
p_{\triangle}^{ER} &= p^6
\end{aligned}
\tag{8}
$$

The triadic random graph model allows us to deviate from this probability distribution. Therefore, we can enhance or suppress certain substructures as compared to ER graphs.

IV. RESULTS

A. Z-score profiles

In order to examine the impact of the triad distribution for the STS on the Z-score profile of the total network we did extensive samplings on the 16-dimensional simplex defined by the probability distributions (7). Samplings were performed for both systems of size 49 and 63. For the computation of the Z-score profiles we used the mfinder software (version 1.2) [45] and averaged the Z-score for each vector $\vec{\mathcal{P}}$ over multiple samples. It shall be noticed that the software only considers those triad configurations which have all three nodes attached to at least one edge. Thus, there are no Z-scores for the subgraphs 1 ($\triangle$), 2 ($\measuredangle$), and 3 ($\measuredangle$) of Fig. 1. Yet, of course it is neccessary to account for them in the input distributions for the STS.

Fig. 5 displays exemplary results obtained from the sampling. Plots a, b, and c show the distributions imposed on the STS (blue circles). This distribution already determines the expectation value for the link-density of the network. E.g. suppose 60% of the Steiner Triples adopt pattern $\wedge$ (which has two of the six possible links being set) and 40% adopt $\triangle$ (five of the
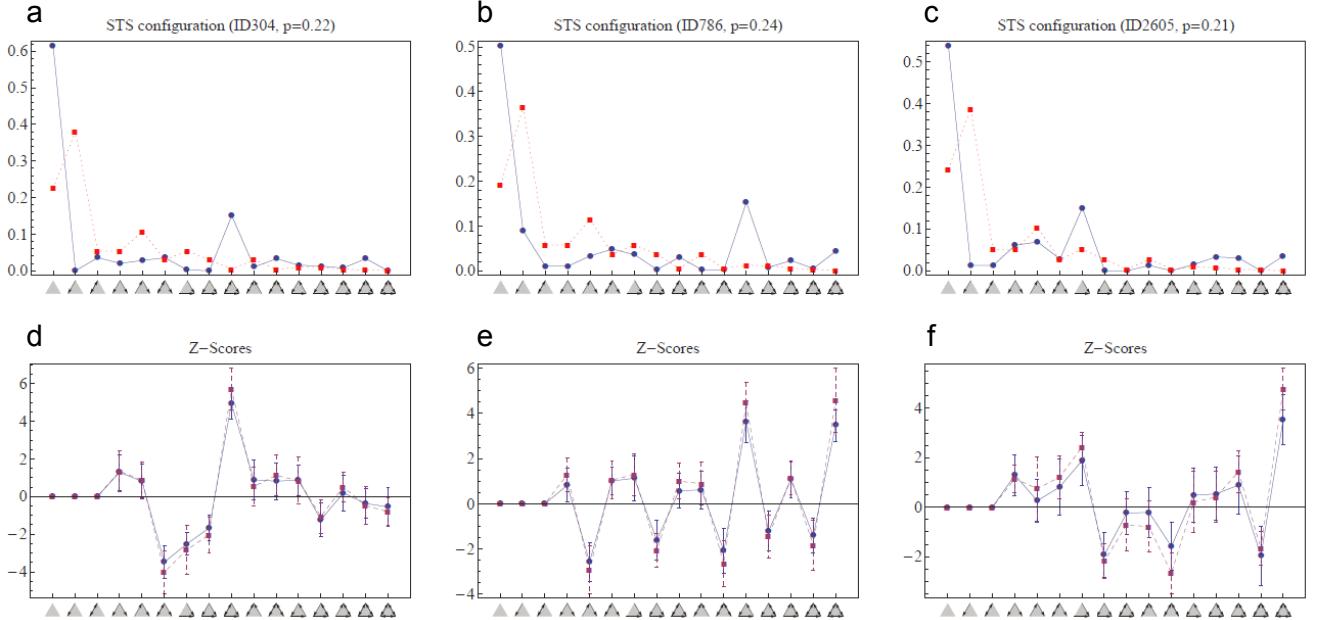
FIG. 5. (Color online) **Top:** Distribution of triad configurations for the Steiner Triples (blue circles) and expected distribution of triad configurations for ER graphs with the same link density (red squares). **Bottom:** Z-scores obtained from networks sampled from the distributions above for systems of size $N = 49$ (blue circles) and $N = 63$ (violet squares), averaged over 15 sample networks.
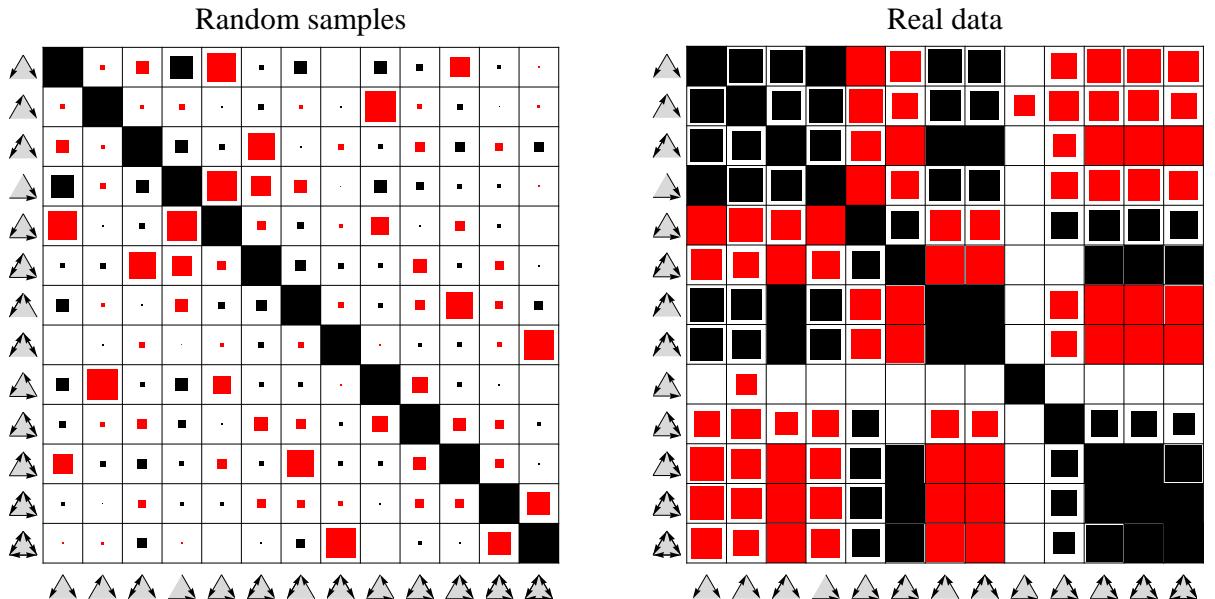

FIG. 6. (Color online) **Left:** Pattern cross correlations in 2×10^4 randomly sampled distributions on Steiner Triple Systems. **Right:** Correlations obtained from real data sets (Table II). The length of the squares indicates the magnitude (0 to 1). Black and red shading corresponds to positive and negative values, respectively. Shown are significant entries at a level of 5%.

six links being set) then the expected density will be $p = (0.6 \cdot 2 + 0.4 \cdot 5)/6 \approx 53\%$. For comparison we also plot the distribution one would expect on the STs for a dyadic ER graph with the same link density as given by Eq. (8) (red squares, dashed line in Fig. 5).

Plots d, e, and f in Fig. 5 show the Z-score profiles obtained from the input distributions above for networks of size 49 (blue circles) and 63 (violet squares). Displayed are the mean values averaged over 15 samples for each distribution. For systems with no higher order structure, such as ER graphs, all Z-scores are expected to vanish. However, for the triadic random graph model, we observe Z-scores with magnitudes larger than five, implying that certain motifs appear five standard deviations more frequently than expected for the randomized ensemble. Thus, triadic random graphs are capable of modelling structure of higher than dyadic order. It shall be emphasized that this higher order structure does not stem from mesoscopic group structure; all Steiner Triples, and therefore all nodes, have the same parameters. In accordance with the literature [23] a larger system size results in a larger magnitude of the Z-scores. However, the shape of the Z-score profiles is size independent.

B. Z-score correlations

For the interpretation of triad significance profiles observed in real networks it is important to be aware of correlations between the Z-scores of pairs of triad patterns, which inherently already arise solely for statistical reasons.

We did extensive uniform sampling of the 16-dimensional simplex spanned by the parameter space of the triadic random graph model (7). In fact, we sampled more than 2×10^4 distinct distributions. For each of the distributions, we generated five network instances and we evaluated the average Z-score profiles. Using the latter, we can evaluate cross correlations between pairs of Z-scores over the input distributions sampled. For two patterns, i and j, it is

$$C_{Z_i, Z_j} = \frac{\langle Z_i\, Z_j \rangle - \langle Z_i \rangle \langle Z_j \rangle}{\sigma_{Z_i}\, \sigma_{Z_j}}. \tag{9}$$

The averages are taken over all sampled STS distributions considered for the evaluation of the correlation matrix. The statistical significance of the correlation is tested by means of a t-test.

Fig. 6 (a) shows the correlation matrix between pairs of Z-scores when sampling randomly. Considered are significant correlations at a level of 5%. The side lengths of the squares indicate the magnitudes of the correlation coefficients between the corresponding subgraphs. Positive values are colored in black, negative ones in red. The magnitude (zero to one) is proportional to the length of the squares. One can clearly see that certain Z-scores are strongly anti-correlated with each other while others are positively correlated. To keep track of the impact of the

link-density on potential correlations, the distributions are grouped in bins of width 0.05. We evaluated seperate correlation matrices for each of the link-density ranges. It turns out that correlations and anti-correlations occur consistently between the same sets of triad patterns for all link-densities sampled.

In order to distinguish between Z-scores which actually describe characteristics of the networks from purely statistical artifacts we also investigated Z-score correlations over various real-world networks. Fig. 6 (b) shows the correlation matrix obtained from the 16 real-world data sets shown in table II. We observe that the most pronounced correlations found in the ensemble of triadic random graphs also appear in the real data sets. This implies that the expressive power of single (anti") motifs might be limited and one should rather consider the Z-score profile in whole. Table III displays the ten strongest cross correlations between pairs of triadic subgraph patterns which were found in our random samples of the triadic random graph ensemble together with the correlation coefficients found in the real data for the respective pairs of triad patterns. Apparently, nine of the top ten (anti") correlations of the statistical data are also found in the real systems. However, not all entries of correlation matrix obtained from the triadic random graphs are reflected in Fig. 6 b): e.g. patterns △ and △ are anti-correlated in the random ensemble, while being strongly positively correlated in the real data. This gives rise to the conjecture that this correlation captures valuable information about the systems' structure. Contrary, e.g. the correlation between patterns △ and △ seems to stem from statistical roots.

Investigations of correlations in the appearance of subgraph motifs have been done before by Ginoza et al. [52]. Yet, their work focuses on correlations within the randomization process of single networks. They consider motifs in two particular networks, namely for the transcriptional regulatory networks of *E. coli* and *S. cerevisiae*. One of their key results is that the abundances of patterns △, △, and △ are strongly mutually correlated, while being anti-correlated with pattern △. Our approach however considers correlations which appear over multiple network instances and is therefore complementary to the one in [52]. Again, Fig. 6 (a), displays our observed correlations between subgraph patterns which occur solely for statistical reasons. In accordance with Ginoza et al. we find strong correlations between patterns △ and △, as well as strong anti-correlation of them with △. However, the former are hardly correlated with pattern △ (in fact, the correlation coefficient is even slightly negative). Although, doubtlessly, in most real networks there is a strong mutual (anti-)correlation in the abundance of subgraphs △, △, △, and △, our results indicate that they do not neccessariliy follow for statistical reasons.

Data set	△	△	△	△	△	△	△	△	△	△	△	△	△
C. Elegans [46–48]	-16.5	-6.29	-24.23	-11.99	12.43	24.48	-27.02	-16.3	-5.02	2.59	27.29	13.15	9.64
Political blogs [46, 49]	-76.09	-51.28	-49.36	-58.19	55.26	40.28	-54.07	-31.17	-2.32	2.97	47.19	27.05	24.82
E. Coli (v. 1.1) [45, 50]	-12.23	-12.23	0	-12.23	12.23	0	0	0	0	0	0	0	0
English book [24, 45]	26.09	13.58	14.35	22.8	-22.52	-10	24.67	13.51	-1.39	-6.59	-21.84	-13.58	-5.53
French book [24, 45]	31.51	26.31	13.4	31.52	-29.1	-10.16	16.17	12.03	-11.5	-12.34	-15.07	-12.33	-4.72
Japanese book [24, 45]	15.01	12.05	13.43	14.97	-14.39	-7.94	12.13	9.27	-4.76	-9.92	-7.4	-8.3	-3.07
Spanish book [24, 45]	26.58	27.5	13.57	23.77	-22.3	-4.16	29.35	12.39	-13.22	-19.82	-25.22	-10.99	-7.57
leader2Inter [24, 45]	-2.25	-1.2	-2.58	-1.22	0.81	1.33	-3.24	-4.5	0.38	1.15	2.31	1.8	3.53
prisonInter [24, 45]	-6.06	-3.71	-10.14	-9.06	4.31	7.84	-8.26	-13.83	0.4	1.99	5.42	7.49	11.93
Electr. circ. (s208) [45]	1.63	-9.57	0	1.63	-1.63	0	0	0	11.01	0	0	0	0
Electr. circ. (s420) [45]	1.61	-17.21	0	1.61	-1.61	0	0	0	20.74	0	0	0	0
S. Cerevisiae [45, 51]	-13.73	-13.52	-0.96	-13.66	13.6	-0.35	-5.91	0	-0.17	9.9	3.94	0	0

TABLE II. Z-scores observed in real-world data sets.

Rank	patterns	random samples	real data
1	△ , △	-0.780487	-0.527285
2	△ , △	-0.74238	-0.977566
3	△ , △	-0.732722	-0.998263
4	△ , △	-0.730034	-0.989075
5	△ , △	-0.661864	-0.985936
6	△ , △	-0.661757	-0.996671
7	△ , △	0.578477	0.990511
8	△ , △	-0.562423	0.958279
9	△ , △	-0.49549	-0.703186
10	△ , △	-0.488278	-0.835307

TABLE III. Top 10 (anti-)correlations between subgraph patterns found in the synthetic random samples, as well as the corresponding correlations observed in real-world data sets.

C. Degree distributions

An important characteristic of complex networks is their degree distribution.

In dyadic ER graphs, the node degrees are expected to be Poissonian distributed:

$$\mathcal{P}\left(k=\kappa\right) = e^{-\langle k \rangle} \frac{\langle k \rangle^{\kappa}}{\kappa!} \tag{10}$$

This holds for both in and out-degrees.

To derive the expected in-degree distribution for triadic random graphs, consider an arbitrary node i. It is part of $(N-1)/2$ Steiner Triples. Now let s_i be a random variable indicating the number of i's Steiner Triples in which a single edge is directed towards it. Further, be d_i the random variable indicating the number of its Steiner Triples with two links directed towards it. From the probabilities in Eq. (7) we can directly infer the probabilities for a single ST to contribute to s_i and d_i, respectively:

$$p\left(s_i\right) = \frac{1}{3}\left[p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)\right] + \frac{2}{3}\left[p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)\right] + p\left(△\right)+p\left(△\right)$$

$$p\left(d_i\right) = \frac{1}{3}\left[p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)+p\left(△\right)\right] + \frac{2}{3}\left[p\left(△\right)+p\left(△\right)\right] + p\left(△\right) \tag{11}$$

Since the model parameters are the same for all nodes, the expectation values for s and d will also be the same for all i:

$$\langle s \rangle = \langle s_i \rangle = \frac{N-1}{2}p\left(s_i\right)$$
$$\langle d \rangle = \langle d_i \rangle = \frac{N-1}{2}p\left(d_i\right) \tag{12}$$

Each of the $(N-1)/2$ Steiner Triples of node i has either no, one, or two edges directed towards it. Therefore, the joint probability distribution of s_i and d_i is given by the multinomial:

$$p\begin{pmatrix} s_i = n_s \\ d_i = n_d \end{pmatrix} = \begin{pmatrix} \frac{N-1}{2} \\ n_s, n_d, \frac{N-1}{2} - n_s - n_d \end{pmatrix} p(s_i)_s^n \, p(d_i)_d^n \, (1 - p(s_i) - p(d_i))^{\frac{N-1}{2} - n_s - n_d}$$

$$= \begin{pmatrix} \frac{N-1}{2} \\ n_s, n_d, \frac{N-1}{2} - n_s - n_d \end{pmatrix} \left(\frac{2}{N-1} \right)^{n_s + n_d} \langle s \rangle^{n_s} \, \langle d \rangle_d^n \left(1 - \frac{2(\langle s \rangle + \langle d \rangle)}{N-1} \right)^{\frac{N-1}{2} - n_s - n_d} \tag{13}$$

For the second equality we used Eq. (12). For large sparse systems and $\langle s \rangle, \langle d \rangle \ll N$ we find

$$\lim_{N \to \infty} p \begin{pmatrix} s_i = n_s \\ d_i = n_d \end{pmatrix} = \lim_{N \to \infty} \frac{\left(\frac{N-1}{2} \right)!}{\left(\frac{N-1}{2} - n_s - n_d \right)!} \left(\frac{2}{N-1} \right)^{n_s + n_d} \frac{\langle s \rangle^{n_s}}{n_s!} \frac{\langle d \rangle^{n_d}}{n_d!}$$

$$\times \underbrace{\left(1 - \frac{\langle s \rangle + \langle d \rangle}{\frac{N-1}{2}} \right)^{\frac{N-1}{2}}}_{\to e^{-\langle s \rangle - \langle d \rangle}} \underbrace{\left(1 - \frac{\langle s \rangle + \langle d \rangle}{\frac{N-1}{2}} \right)^{-(n_s + n_d)}}_{\to 1}$$

$$= \lim_{N \to \infty} \underbrace{\frac{\frac{N-1}{2} + \frac{N-3}{2} + ... + \frac{N-1}{2} - n_s - n_d + 1}{\left(\frac{N-1}{2} \right)^{n_s + n_d}}}_{\to 1} \frac{\langle s \rangle^{n_s}}{n_s!} \frac{\langle d \rangle^{n_d}}{n_d!} e^{-\langle s \rangle - \langle d \rangle} \tag{14}$$

$$= \frac{\langle s \rangle^{n_s}}{n_s!} \frac{\langle d \rangle^{n_d}}{n_d!} e^{-\langle s \rangle - \langle d \rangle}.$$

The in-degree of node i is

$$k_i^{\text{in}} = s_i + 2d_i. \tag{15}$$

The probability distribution for node i to have in-degree κ is thus

$$p\left(k^{\text{in}} = \kappa\right) = \sum_{n_s=0}^{\frac{N-1}{2}} \sum_{n_d=0}^{\frac{N-1}{2}} p \begin{pmatrix} s_i = n_s \\ d_i = n_d \end{pmatrix} \delta_{\kappa, n_s + 2n_d}$$

$$= e^{-\langle s \rangle - \langle d \rangle} \sum_{n_d=0}^{\frac{\kappa}{2}} \frac{\langle s \rangle^{\kappa - 2 n_d}}{(\kappa - 2n_d)!} \frac{\langle d \rangle^{n_d}}{n_d!} \tag{16}$$

where δ is the Kronecker delta ($\delta_{i,j} = 1$ if $i = j$, 0 otherwise). In the limit $\langle d \rangle \to 0$ the distribution is Poissonian. With $\langle d \rangle$ approaching $\frac{1}{2} \langle k^{in} \rangle$, the distribution becomes broader, implying larger deviations from $\langle k \rangle$. Fig. 7 shows distributions of Eq. (16) with fixed $\langle k \rangle = \langle s \rangle + 2 \langle d \rangle = 100$ for various ratios of $r = \langle s \rangle / \langle d \rangle$ together with the corresponding Poissonian.

The out-degree distribution can be derived analogously. In this case, only the probabilites for the triads with a single out-going edge, $p(s_i^{\text{out}})$, and for two out-going edges, $p(d_i^{\text{out}})$, need to be adjusted accordingly.

$$p\left(s_i^{\text{out}}\right) = \frac{1}{3} \left[p(\angle) + p(\wedge) + p(\triangle) + p(\triangle) \right] + \frac{2}{3} \left[p(\angle) + p(\wedge) + p(\triangle) + p(\wedge) + p(\triangle) + p(\triangle) \right] + p(\wedge) + p(\triangle)$$

$$p\left(d_i^{\text{out}}\right) = \frac{1}{3} \left[p(\wedge) + p(\wedge) + p(\triangle) + p(\wedge) + p(\triangle) + p(\triangle) \right] + \frac{2}{3} \left[p(\triangle) + p(\triangle) \right] + p(\triangle) \tag{17}$$

D. Design of significance profiles

To achieve the goal of designing networks with certain triad significance profiles, it is important to understand the relationship between the distribution of triad con-figurations on Steiner Triples and the Z-scores obtained from their ensembles. Therefore, we also investigated the cross correlations between the ST configurations and the
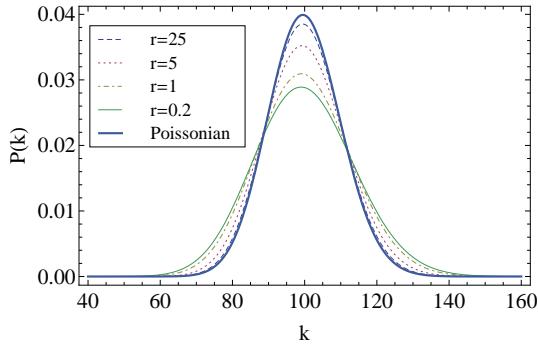
FIG. 7. (Color online) Degree distributions for mean degree $\langle k \rangle = 100$ and various ratios $r = \langle s \rangle / \langle d \rangle$.
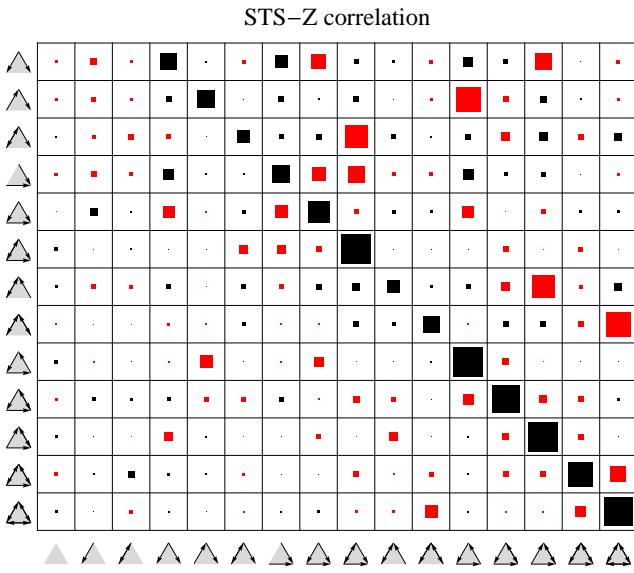

FIG. 8. (Color online) Correlation matrix between triad configurations on ST and the resulting Z-score profiles obtained from 5000 configurations. The length of the squares indicates the magnitude (0 to 1). Black and red shading corresponds to positive and negative values, respectively.

obtained corresponding Z-scores:

$$\widetilde{C}_{\mathcal{P}_i, Z_j} = \frac{\langle \mathcal{P}_i Z_j \rangle - \langle \mathcal{P}_i \rangle \langle Z_j \rangle}{\sigma_{\mathcal{P}_i} \sigma_{Z_j}}. \tag{18}$$

Results are presented in Fig. 8. Of course, there is a strong correlation between the imposed triad patterns on the Steiner Triples and the Z-scores of these patterns. However, as for the Z-score-Z-score cross correlations, again we observe strong anti-correlations between certain patterns. As before, the observations are valid for all examined link densities. Correlations between the input distributions on the STS and the obtained over-all Z-score profiles can be helpful in designing systems with pre-defined significance profiles.

For a simplistic approach, we assume a linear relation
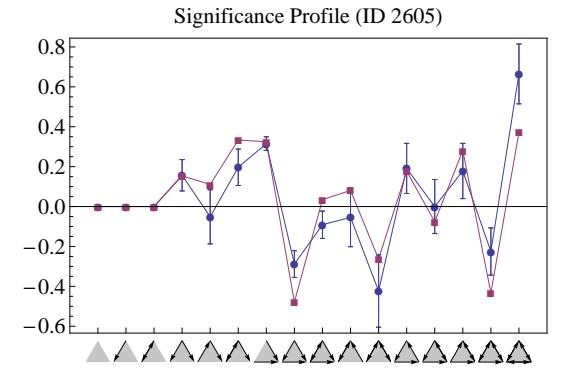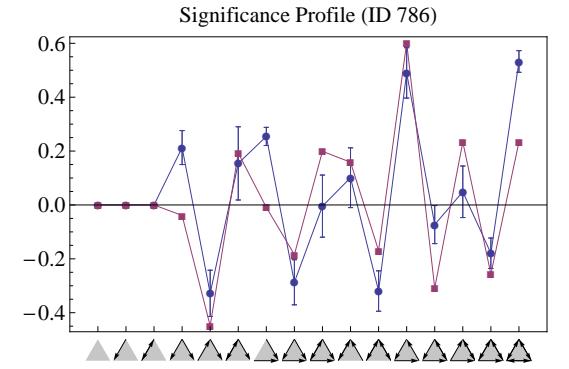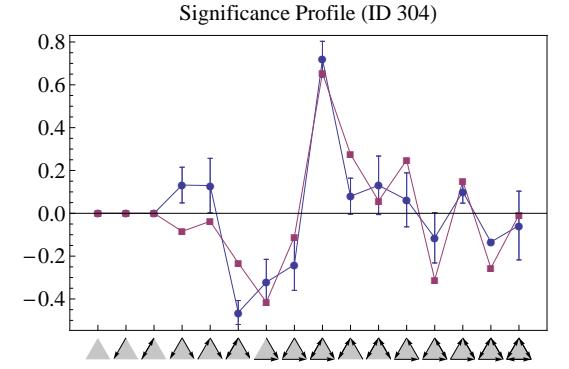
FIG. 9. (Color online) Significance profiles corresponding to the Z-scores in Fig. 5 for systems of size 49 (blue circles). The violet squares indicate the prediction obtained from the input distribution $\vec{\mathcal{P}}$ by assuming $\vec{\text{SP}} \propto \widetilde{C} \vec{\mathcal{P}}$.

between the input distribution $\vec{\mathcal{P}}$ and the significance profile, conveyed by the correlation matrix $\widetilde{C}$ (Fig. 8),

$$\vec{\text{SP}} \propto \widetilde{C} \vec{\mathcal{P}}. \tag{19}$$

In order to design systems with pre-defined significance profiles, it is neccessary to map the latter to a corresponding input distribution, which can be realized by means of the pseudo inverse matrix $\widetilde{C}^{-1}$.

$$\vec{\mathcal{P}} \propto \widetilde{C}^{-1} \vec{\text{SP}} \tag{20}$$
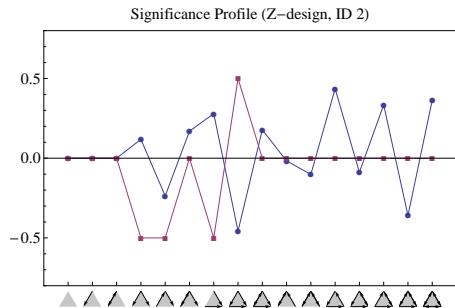
FIG. 10. (Color online) Attempt to model the significance profile indicated by the violet squares. However, networks with their parametrization obtained from Eq. (20) yield very different significance profiles.

Fig. 9 shows the significance profiles corresponding to the Z-scores in Fig. 5 together with the prediction obtained from Eq. (19). The predictions agree very well with the actually observed profiles. However, attempts to model arbitrary significance profile with the linear relation Eq. (20) will not succeed in all cases as shown in Fig. 10.

This may be for various reasons. On the one hand the relationship between $\vec{\mathcal{P}}$ and the significance profile is certainly not entirely linear. Secondly, not all significance profiles are necessarily realizable, e.g. think of a SP with all patterns being overrepresented. Furthermore, the triadic random graph model describes the most simplistic model based on STS, which, e.g. does not account for individual node properties. This is also reflected in the fact that the degree distributions of triadic random graphs are close to a Poissonian. A formulation of more specific models based on STS may overcome these shortcomings. Still, these first steps open the way to efficiently generate networks in which certain motifs are over- or underrepresented and thus enable systematic investigations of the functional significance of these motifs.

V. CONCLUSIONS

Over the last decade the over- and underrepresentation of particular sub-network patterns has attracted high attention. This led to the hypothesis that, instead of links, they serve as the building blocks of network structures [23]. The fact that only a small portion of triad configurations can actually be specified independently poses a challange to the formulation of generative models which account for higher order substructures. Based on sets of pair-disjoint triads, so called Steiner Triple Systems, we have introduced a novel class of generative models. The most simple realilzation of such models assumes the same probability distribution over the possible triad patterns for all Steiner Triples in the system. We referred to them as triadic random graph models. By extensive samplings we proved that, in constrast to ER graphs, even this most simplistic model is capable of inducing non-vanishing Z-scores. Furthermore, we discovered inevitable correlations between triad patterns whith respect to their abundance. These occur solely for statistical reasons. This dependence in the appearance of subgraph patterns should be taken into account when attributing functional relevance to network motifs. Moreover, we unveiled correlations between the probability distributions on the Steiner Triples and the observed Z-score profiles over the whole network. These are helpful for designing ensembles of networks with pre-defined significance profiles which can facilitate a systematic study of the effect of motif distributions on network dynamics. Finally, we could also calculate the degree distributions of triadic random graphs analytically. We found it to be similar, yet not identical to a Poissionian. Depending on the input distribution $\vec{\mathcal{P}}$, the degree distribution is broader than a Poissonian.

The triadic random graph model assumes all nodes to be equal and thus can be considered the triadic analogon to ER graphs. However, in many real-world systems, individual node properties like the popularity or activity of vertices play a crucial role. Future models based on Steiner Triple Systems which, e.g. aim to predict hitherto undiscovered links may include those parameters in Eq. (6). In addition, the introduced framework also allows for the definition of models for signed networks, *i.e.* graphs with positive or negative edges which play in important role in the social sciences in the context of structural balance theory [53, 54] as well as in the biosciences where they are used to model excitatory and inhibitory links in neural or gene-regulation networks [55].

[1] P. Erdös and A. Rényi, Publicationes Mathematicae (Debrecen) **6**, 290 (1959).
[2] P. Erdös and A. Rényi, Evolution **5**, 17 (1960).
[3] A. Barabási, Science **286**, 509 (1999).
[4] R. Albert and A.-L. Barabási, Reviews of Modern Physics **74**, 47 (2002).
[5] P. Krapivsky and S. Redner, in *Statistical Mechanics of Complex Networks*, Lecture Notes in Physics, Vol. 625, edited by R. Pastor-Satorras, M. Rubi, and A. Diaz-Guilera (Springer Berlin Heidelberg, 2003) pp. 3–22.
[6] S. Dorogovtsev and J. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (2003).
[7] M. E. Newman, SIAM review **45**, 167 (2003).
[8] I. Lodato, S. Boccaletti, and V. Latora, Europhysics Letters (EPL) **78**, 28001 (2007).
[9] J. Besag and R. Statistical, Society **36**, 192 (1974).
[10] D. Strauss, SIAM Review **28**, 513 (1986).
[11] P. W. Holland and S. Leinhardt, Journal of the American Statistical Association **76**, 33 (1981).
[12] S. Wasserman and P. Pattison, Psychometrika **61**, 401

(1996).

[13] T. A. B. Snijders, Journal of Social Structure **3**, 1 (2002).

[14] M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris, Journal Of Statistical Software **24**, 1 (2008).

[15] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, Journal Of Statistical Software **24**, nihpa54860 (2008).

[16] M. Molloy and B. Reed, Random Structures & Algorithms **6**, 161 (1995).

[17] M. Molloy and B. Reed, Combinatorics, Probability and Computing **7**, 295.

[18] P. W. Holland, K. B. Laskey, and S. Leinhardt, Social Networks **5**, 109 (1983).

[19] K. Nowicki and T. A. B. Snijders, Journal of the American Statistical Association **96**, 1077 (2001).

[20] T. A. B. Snijders and K. Nowicki, Journal of Classification **14**, 75 (1997).

[21] Y. J. Wang and G. Y. Wong, Journal of the American Statistical Association **82**, 8 (1987).

[22] B. Karrer and M. E. J. Newman, Physical Review E - Statistical, Nonlinear and Soft Matter Physics Phys. Rev. E, Stat. Nonlinear Soft Matter Phys. (USA), **83**, 016107 (2010).

[23] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science (New York, N.Y.) **298**, 824 (2002).

[24] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science (New York, N.Y.) **303**, 1538 (2004).

[25] O. Sporns and R. Kötter, PLoS Biology **2**, e369 (2004).

[26] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nature genetics **31**, 64 (2002).

[27] I. Albert and R. Albert, Bioinformatics (Oxford, England) **20**, 3346 (2004).

[28] Y.-H. Eom, S. Lee, and H. Jeong, Journal of theoretical biology **241**, 823 (2006).

[29] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, Journal of the Royal Society, Interface / the Royal Society **7**, 1093 (2010).

[30] S. Sakata, Y. Komatsu, and T. Yamamori, Neuroscience research **51**, 309 (2005).

[31] M. E. J. Newman, Physical Review Letters **103**, 5 (2009).

[32] E. M. Jin, M. Girvan, and M. E. J. Newman, Physical Review E **6402**, 046132 (2001).

[33] P. Holme and B. J. Kim, Physical Review E **65**, 2 (2001).

[34] K. Klemm and V. Eguíluz, Physical Review E **65**, 1 (2002).

[35] S. Bansal, S. Khandelwal, and L. A. Meyers, Computational Complexity (2008).

[36] M. Angeles Serrano and M. Boguñá, Physical Review E - Statistical, Nonlinear and Soft Matter Physics **72**, 036133 (2005).

[37] J. Park and M. E. J. Newman, Physical Review E **72**, 1 (2005), arXiv:0412579 [cond-mat].

[38] J. Reichardt, R. Alamino, and D. Saad, PLoS ONE **6**, e21282 (2011).

[39] M. E. Beber, C. Fretter, S. Jain, N. Sonnenschein, M. Müller-Hannemann, and M.-T. Hütt, Journal of the Royal Society, Interface / the Royal Society **9**, 3426 (2012).

[40] C. Fretter, M. Müller-Hannemann, and M.-T. Hütt, (2012).

[41] Steiner triple systems are a special case of the more general $t(v, k, \lambda)$ or $S_\lambda(t, k, v)$ designs, where v denotes the number of points and k denotes the cardinality of the blocks (three for triangles). For any set T of t points, there are exactly λ blocks incident with all points in T. Thus, Steiner triple systems are $2(v, 3, 1)$ or $S_1(2, 3, v)$. For more details see e.g. [43].

[42] M. J. Hall, *Combinatorial theory* (John Wiley & Sons, 1986).

[43] J. H. Van Lint and R. M. Watson, *A Course in Combinatorics* (Cambridge University Press, 1992) Chap. 1.

[44] T. P. Kirkman, The Cambridge and Dublin Mathematical Journal **2**, 191 (1847).

[45] U. Alon, "http://www.weizmann.ac.il/mcb/urialon/," (2012).

[46] M. E. J. Newman, "http://www-personal.umich.edu/~mejn/netdata/," (2012).

[47] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[48] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, Philosophical Transactions of the Royal Society B: Biological Sciences **314**, 1 (1986).

[49] L. A. Adamic and N. Glance, in *Proceedings of the 3rd international workshop on Link discovery*, LinkKDD '05 (ACM, 2005) pp. 36–43.

[50] S. Mangan and U. Alon, Proceedings of the National Academy of Sciences of the United States of America **100**, 11980 (2003).

[51] M. C. Costanzo, Nucleic Acids Research **29**, 75 (2001).

[52] R. Ginoza and A. Mugler, Physical Review E **82**, 1 (2010).

[53] F. Heider, Psychological Review **51**, 358 (1944).

[54] D. Cartwright and F. Harary, Psychological review **63**, 277 (1956).

[55] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, Proceedings of the National Academy of Sciences of the United States of America **101**, 4781 (2004).